
Preface

The primary goal of this book is to present to the scientific and management communities a selection of applications using more recent Soft Computing (SC) and Computing with Words and Perceptions (CWP) models and techniques meant to solve the economics and financial problems. The selected examples could also serve as a starting point or as an opening out, in the SC and CWP techniques application to a wider range of problems in economics and finance.

Decision making in the present world is becoming more and more sophisticated, time consuming and difficult for human beings who require more and more computational support. This book addresses the significant increase on research and applications of Soft Computing and Computing with Words and Perceptions for decision making in Economics and Finance in recent years. Decision making is heavily based on information and knowledge usually extracted from the analysis of large amounts of data. Data mining techniques enabled with the capability to integrate human experience could be used for a more realistic business decision support. Computing with Words and Perceptions introduced by Lotfi Zadeh, can serve as a basis for such extension of traditional data mining and decision making systems. Fuzzy logic as a main constituent of CWP gives powerful tools for modeling and processing linguistic information defined on numerical domain.

Decision making techniques based on fuzzy logics in many cases have demonstrated better performance than competing approaches. The reason is that traditional, bivalent-logic-based approaches, are not a good fit to reality — the reality of pervasive imprecision, uncertainty and partiality of truth. On the other hand, traditional probabilistic interpretation of uncertainties in practice does not always correspond to the nature of uncertainties that often appear as the effects of subjective estimations. The list of practical situations, when it seems better to avoid the traditional probabilistic interpretation of uncertainty is very long. The centrepiece of fuzzy logic that everything is, or is allowed to be, a matter of degree, makes it possible to better deal with perception-based information. Such information plays an essential role in economics, finance and, more generally in all domains in which human perceptions and emotions are in evidence. For instance, it is the case for the studies of the capital markets/financial engineering including financial time series modeling; price projections for stocks, volatility analysis and the pricing of options and derivatives; and risk management to mention few.

The book consists of two parts: Data Mining and Decision Making. An introductory chapter by Lotfi A. Zadeh called “Precisiated Natural Language” describes the conceptual structure of Precisiated Natural Language (PNL), which

can be employed as a basis for computation with perceptions. PNL attempts to make possible treating propositions drawn from a natural language as objects of computation, capturing two fundamental facets of human cognition (a) partiality (of understanding, truth, possibility, etc.) and (b) granularity (clumping of values of attributes and forming granules with words as labels). The chapter shows that PNL has much higher expressive power than existing approaches to natural language processing based on bivalent logic. Its high expressiveness is based on the concept of a generalized constraint, which represents the meaning of propositions drawn from a natural language capturing their partiality and granularity. This chapter settles the conceptual basis for the rest of the book.

The first part of the book presents novel techniques of Data Mining. Researchers in the data mining field have traditionally focused their efforts on obtaining algorithms in order to deal with huge amounts of data. It is nevertheless true that the results we obtain using these algorithms are of limited use in practice hence limiting data mining spread use and acceptance in many real-world situations. One of the reasons for this is that purely statistical approaches, which do not consider the experience of the experts do not hit in the actual problem. Perception-based data mining should be able to manipulate with linguistic information, fuzzy concepts and perception based patterns of time series. That is why, additionally to classic techniques of data mining and classification algorithms such as decision trees or Bayesian classifiers, the subsequent chapters study other data mining operations such as clustering, moving approximations, fuzzy association rule generation more suitable to work with perceptual patterns. The subsequent chapters describe novel techniques of perception-based data mining with their applications to typical economics and finance problems.

The chapter titled “Towards Human-Consistent Data-Driven Decision Support Systems via Fuzzy Linguistic Data Summaries” by Janusz Kacprzyk and Sławomir Zadrozny focuses on construction of linguistic summaries of data. Summarization as one of the typical tasks of data mining provides efficient and human consistent means for the analysis of large amounts of data to be used for a more realistic business decision support. The chapter shows how to embed data summarization within the fuzzy querying environment, for an effective and efficient computer implementation. Realization of Zadeh’s computing with words and perception paradigm through fuzzy linguistic database summaries, and indirectly fuzzy querying, can open new vistas in data driven and also, to some extent, in knowledge driven and Web-based Decision Support System.

The next chapter titled “Moving Approximation Transform and Local Trend Associations in Time Series Data Bases” by Batyrshin I., Herrera-Avelar R., Sheremetov L., and Panova A. describes a new technique of time series analysis based on a replacement of time series by the sequences of slopes of linear functions approximating time series in sliding windows. Based on Moving Approximation (MAP) Transform several measures of local trend associations can be introduced which are invariant under linear transformation of time series. Due to this very important property the local trend association measures can serve as basic measures of similarity of time series and time series patterns in most of problems of time series data mining. The chapter considers several examples of

application of local trend association measures to construction of association networks for systems described by time series data bases. MAP can be used also as a basis for definition of perception based trend patterns like “quickly increasing”, “very slowly decreasing” etc in intelligent decision making systems including as a part expert knowledge and time series data bases. Nowadays Discrete Fourier Transform is a main technique for analysis of time series which describe signal propagation and oscillating processes, when the concept of frequency plays a key role. MAP transform can serve as a main instrument for analysis of local trends and tendencies of non-oscillating processes which is important for economic and financial applications.

Most of information used in economics and finance is stored in time series; that is why, the book pays special attention to time series data mining (TSDM). Development of intelligent question answering systems supporting decision making procedures related with time series data bases needs to formalize human perceptions about time, time series values, patterns and shapes, about associations between patterns and time series, etc. The chapter called “Perception Based Patterns in Time Series Data Mining” by Batyrshin I., Sheremetov L., and Herrera-Avelar R. presents an overview of the current techniques and analyse them from the point of view of their contribution to perception based TSDM. The survey considers different approaches to description of perception based patterns which use sign of derivatives, scaling of trends and shapes, linguistic interpretation of patterns obtained as result of clustering, a grammar for generation of complex patterns from shape primitives, temporal relations between patterns. These descriptions can be extended by using fuzzy granulation of time series patterns to make them more adequate to perceptions used in human reasoning. Several approaches to relate linguistic descriptions of experts with automatically generated texts of summaries and linguistic forecasts are considered. Finally, it is discussed the role of perception based time series data mining and computing with words and perceptions in construction of intelligent systems that use expert knowledge and decision making procedures in time series data base domains.

The next chapter titled “Perception Based Functions in Qualitative Forecasting” by Batyrshin I. and Sheremetov L. discusses application of fuzzy perception based functions (PBF) to qualitative forecasting of a new product life cycle. PBF are given by a sequence of rules R_k : *If T is T_k then S is S_k* , where T_k are perception based intervals defined on the domain of independent variable T , and S_k are perception based shape patterns of variable S on interval T_k . Intervals T_k can be expressed by words like *Between N and M , Approximately M , Middle of the Day, End of the Week* etc. Shape patterns S_k can be expressed linguistically, e.g. as follows: *Very Large, Increasing, Quickly Decreasing and Slightly Concave* etc. The authors consider new parametric patterns used for modeling convex-concave shapes of PBF and propose a method of reconstruction of PBF with these shape patterns. These patterns can be used also for time series segmentation in perception based time series data mining.

The chapter titled “Towards Automated Share Investment System” by Dymitr Ruta describes a classification model that learns the transaction patterns from optimally labelled historical data presented as time series and accordingly gives

VIII Preface

the profit-driven decision for the current-day transaction. Contrasting to the traditional regression-based approaches, a proposed model can facilitate the job of a busy investor preferring a simple decision on the current day transaction: buy, wait, sell that would maximise his return from the investment. The model is embedded into an automated client-server platform which automatically handles data collection and maintains client models on the database. The prototype of the system was tested over 20 years of NYSE:CSC share price history showing substantial improvement of the long-term profit compared to a passive long-term investment.

The Decision Tree is one of the most popular classification algorithms in current use in Data Mining. The chapter called “Estimating Classification Uncertainty of Bayesian Decision Tree Technique on Financial Data” by Shetinin et al. studies interpretability issues of classification models, which is crucial for experts responsible for making reliable classifications. Decision Trees classification model is combined with the Bayesian model averaging all possible DTs achieving thus the required diversity of the DT ensemble. The authors explore the classification uncertainty of the Bayesian Markov Chain Monte Carlo techniques on some datasets from the StatLog Repository and real financial data. The classification uncertainty is compared within an Uncertainty Envelope technique dealing with the class posterior distribution and a given confidence probability. This technique provides realistic estimates of the classification uncertainty, which can be easily interpreted in statistical terms with the aim of risk evaluation.

Another important method frequently used in data mining is Cluster Analysis. In the chapter “Invariant Hierarchical Clustering Schemes” by Batyrshin I. and Rudas T., the properties of general scheme of parametric invariant clustering procedures based on transformation of proximity function into fuzzy equivalence relation are studied. The scheme gives possibility to build clustering procedures invariant to numeration of objects and to monotone transformation of proximity values between objects. Several examples are used to illustrate the application of proposed clustering procedures to analysis of similarity structures of data.

The second part of the book focuses on the problems of perceptual Decision Making in Economics and Finance. As shown in the previous chapters, the time series analysis has three goals: forecasting (also called predicting), modeling, and characterization. Almost all managerial decisions are based on forecasting and modelling. The ability to model and perform decision modeling and analysis is an essential feature of many real-world applications.

The second part opens with the chapter called “Fuzzy Components of Cooperative Markets” by Milan Mareš deals with the Walras equilibrium model and its cooperative modification also analyzing some possibilities of its fuzzification. The main attention is focused on the vagueness of utility functions of prices, which can be considered for most subjective (utilities) or most unpredictable (prices) components of the model. The elementary properties of the fuzzified model are presented and the adequacy of the suggested fuzzy set theoretical methods to the specific properties of real market models is briefly discussed.

The chapter titled “Possibilistic-Probabilistic Models and Methods of Portfolio Optimization” by Alexander Yazenin, considers a generalization of Markowitz models with fuzzy random variables characterized by corresponding possibility distributions. Such situation is incident for financial assets, particularly, of Russian market. In this case the financial asset profitability can be represented by fuzzy random variable. Approaches to definition of numerical characteristics of fuzzy random variables are analyzed and proposed. Appropriate methods of calculation, in particular within the framework of shift-scaled representation are obtained. Principles of decision making in fuzzy random environment using these possibilistic-probabilistic models are formulated.

In the chapter “Towards Graded and Nongraded Variants of Stochastic Dominance” by Bernard De Baets and Hans De Meyer, a pairwise comparison method for random variables is established. This comparison results in a probabilistic relation on a given set of random variables. The transitivity of this probabilistic relation is investigated, which allows identifying appropriate strict or weak thresholds, depending upon the copula involved, turning the probabilistic relation into a strict order relation. The proposed method can also be seen as a way of generating graded as well as non-graded variants of the concept of stochastic dominance.

The chapter titled “Option Pricing in the Presence of Uncertainty” by Silvia Muzzioli and Huguette Reynaerts studies the derivation of the European option price in the Cox Ross Rubinstein (1979) binomial model in the presence of uncertainty in the volatility of the underlying asset. Two different approaches to the issue that concentrate on the fuzzification of one or both the two jump factors are proposed. The first approach is derived by assuming that both the jump factors are represented by triangular fuzzy numbers. The second approach to the option pricing problem is derived under the assumption that only the up jump factor is uncertain.

In the chapter titled “Non-Stochastic-Model Based Finance Engineering” by Toshihiro Kaino and Kaoru Hirota, a new corporate evaluation model and option pricing model based on fuzzy measures are proposed, which deal with the ambiguous subjectivity evaluation of human in the real world.

The chapter called “Collective Intelligence in Multiagent Systems: Interbank Payment Systems Application” by Luis Rocha-Mier, Leonid Sheremetov and Francisco Villarreal describes a new approach to the interbank net settlement payment systems (NSPS) modeling in order to analyze the actions of individual depositors. The model is developed within the framework of a Collective Intelligence (COIN). This framework is focused on the interactions at the local and the global levels among the consumer-agents (without a global knowledge of the environment model) in order to assure the optimization of the global utility function (GUF). A COIN is defined as a large Multi-Agent System (MAS) with no centralized control and communication, but where there is a global task to complete. Reinforcement learning algorithms are used at the local level, while techniques based on the COIN theory are used to optimize the global behavior. The proposed framework was implemented using Netlogo (agent-based parallel modeling and simulation environment). The results demonstrate that the inter-

bank NSPS is a good experimental field for the application of the COIN theory and shows how the consumer-agents behavior converges to the Nash equilibrium adapting their actions to optimize the GUF.

Finally, the chapter titled “Fuzzy Models in Credit Risk Analysis” by Antonio Carlos Pinto Dias Alves, presents some concepts guiding credit risk analysis using fuzzy logic systems. Fuzzy quantification theory is used to make a kind of multivariate analysis giving more usable answers than traditional Logit or Probit analysis. To make the analysis some interesting accounting indicators are used that can efficiently point the financial health of a company.

At the moment, most of the models in the field of finance engineering are based on the stochastic theory. As shown in this book, the prediction based on these models sometime does not hit in the actual problem. One of the reasons is that they assume a known probability distribution and try to describe a system with the precision obtained by means of the exact probability densities. Nevertheless, as shown in chapter 4, for example, classification model that learns the transaction patterns applying a regression model to support or even make investment decisions is inappropriate as on top of being uncertain and unnecessarily complex requires lots of investor attention and further analysis to make an investment decision. In many cases it is impossible or unnecessary to describe the behavior of modeled system with the precision that may be obtained by means of the exact probability densities. SC models and decision making procedures described in this book, overcome these drawbacks which makes them more suitable for real-world situations. Perception based models can be a powerful tool helping to develop a new generation of human consistent, natural language based and easy to use decision support systems.

Finally, we would like to thank all the contributing authors for excellent research papers, their institutions for the support provided for the research work in the field of soft computing, Studies in Fuzziness and Soft Computing Series Editorial Board and Springer Verlag for their interest in the topic and giving the possibility to publish this book.

March 2006

I. Batyrshin
J. Kacprzyk
L. Sheremetov
L.A. Zadeh

Perception-based Data Mining and Decision Making in
Economics and Finance

Batyrshin, I.; Sheremetov, L.; Zadeh, L.A. (Eds.)

2007, XVI, 367 p., Hardcover

ISBN: 978-3-540-36244-9